

=====

Sequence Listing was accepted.

If you need help call the Patent Electronic Business Center at (866)  
217-9197 (toll free).

Reviewer: Anne Corrigan

Timestamp: [year=2008; month=12; day=10; hr=8; min=36; sec=12; ms=270; ]

=====

Application No: 10591029 Version No: 1.0

**Input Set:****Output Set:**

**Started:** 2008-12-09 15:33:17.645  
**Finished:** 2008-12-09 15:33:20.383  
**Elapsed:** 0 hr(s) 0 min(s) 2 sec(s) 738 ms  
**Total Warnings:** 49  
**Total Errors:** 0  
**No. of SeqIDs Defined:** 49  
**Actual SeqID Count:** 49

Error code	Error Description
W 213	Artificial or Unknown found in <213> in SEQ ID (1)
W 213	Artificial or Unknown found in <213> in SEQ ID (2)
W 213	Artificial or Unknown found in <213> in SEQ ID (3)
W 213	Artificial or Unknown found in <213> in SEQ ID (4)
W 213	Artificial or Unknown found in <213> in SEQ ID (5)
W 213	Artificial or Unknown found in <213> in SEQ ID (6)
W 213	Artificial or Unknown found in <213> in SEQ ID (7)
W 213	Artificial or Unknown found in <213> in SEQ ID (8)
W 213	Artificial or Unknown found in <213> in SEQ ID (9)
W 213	Artificial or Unknown found in <213> in SEQ ID (10)
W 213	Artificial or Unknown found in <213> in SEQ ID (11)
W 213	Artificial or Unknown found in <213> in SEQ ID (12)
W 213	Artificial or Unknown found in <213> in SEQ ID (13)
W 213	Artificial or Unknown found in <213> in SEQ ID (14)
W 213	Artificial or Unknown found in <213> in SEQ ID (15)
W 213	Artificial or Unknown found in <213> in SEQ ID (16)
W 213	Artificial or Unknown found in <213> in SEQ ID (17)
W 213	Artificial or Unknown found in <213> in SEQ ID (18)
W 213	Artificial or Unknown found in <213> in SEQ ID (19)
W 213	Artificial or Unknown found in <213> in SEQ ID (20)

**Input Set:**

**Output Set:**

**Started:** 2008-12-09 15:33:17.645  
**Finished:** 2008-12-09 15:33:20.383  
**Elapsed:** 0 hr(s) 0 min(s) 2 sec(s) 738 ms  
**Total Warnings:** 49  
**Total Errors:** 0  
**No. of SeqIDs Defined:** 49  
**Actual SeqID Count:** 49

Error code

Error Description

This error has occurred more than 20 times, will not be displayed

# SEQUENCE LISTING

<110> WOOD, DAVID W.  
HSII, JUDY  
OAK, SEACHOL  
CONTRERAS, LYDIA  
CHESTNUT, JOHN

<120> SELF-CLEAVING AFFINITY TAGS AND METHODS OF USE

<130> 331772-00103

<140> 10591029

<141> 2008-12-09

<150> PCT/US05/05763

<151> 2005-02-24

<150> 60/548,092

<151> 2004-02-27

<160> 49

<170> PatentIn version 3.3

<210> 1

<211> 504

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
mini-intein derived from the full-length Mycobacterium  
tuberculosis recA intein sequence

<400> 1

gccctcgcag agggcactcg gatcttcgat ccggtcaccg gtacaacgca tcgcatcgag	60
gatgttgctcg gtggggcgcaa gcctattcat gtcgtggctg ctgccaagga cggaacgctg	120
catgcgcggc ccgtggtgctc ctggttcgac cagggaacgc gggatgtgat cggggttcg	180
atcgccggtg ggcgccatcct gtggggcgaca cccgatcaca aggtgctgac agagtacggc	240
tggcgtgccg ccgggggaact ccgcaaggga gacagggtgg cgcaaccgcg acgcttcgat	300
ggattcgggtg acagtgcgcc gattccggcg cgcgtgcagg cgctcgcgga tgccctggat	360
gacaaattcc tgcacgacat gctggcgga gaactccgct attccgtgat ccgagaagtg	420
ctgccaacgc ggcgggcacg aacgttcggc ctcgaggtcg aggaactgca caccctcgtc	480
gccgaagggg ttgttgtaca caac	504

<210> 2

<211> 168

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
mini-intein derived from the full-length Mycobacterium  
tuberculosis recA intein sequence

<400> 2

Ala Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr  
1                  5                  10                  15

His Arg Ile Glu Asp Val Val Gly Gly Arg Lys Pro Ile His Val Val  
                  20                  25                  30

Ala Ala Ala Lys Asp Gly Thr Leu His Ala Arg Pro Val Val Ser Trp  
                  35                  40                  45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly  
                  50                  55                  60

Ala Ile Leu Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly  
65                  70                  75                  80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gln Pro  
                  85                  90                  95

Arg Arg Phe Asp Gly Phe Gly Asp Ser Ala Pro Ile Pro Ala Arg Val  
                  100                  105                  110

Gln Ala Leu Ala Asp Ala Leu Asp Asp Lys Phe Leu His Asp Met Leu  
                  115                  120                  125

Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val Leu Pro Thr Arg  
                  130                  135                  140

Arg Ala Arg Thr Phe Gly Leu Glu Val Glu Glu Leu His Thr Leu Val  
145                  150                  155                  160

Ala Glu Gly Val Val Val His Asn  
                  165

<210> 3

<211> 504

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
TOPO modified intein sequence from SEQ ID NO: 1

<400> 3

```
gccctcgcag agggcactcg gatcttcgat ccggtcacccg gtacaacgca tcgcatcgag      60
gatgttgctg gtggggcgcaa gcctattcat gtcgtggctg ctgccaagga cggaacgctg    120
catgcgcggc ccgtggtgtc ctggttcgac cagggaacgc gggatgtgat cggggtgcgg      180
atcgccggtg gcgccatcct gtggggcgaca cccgatcaca aggtgctgac agagtacggc     240
tggcgtgccg ccgggggaact ccgcaaggga gacagggtgg cgcaaccgcg acgcttcgat     300
ggattcgggtg acagtgcgcc gattccggcg cgcgtgcagg cgctcgcgga tgccctggat     360
gacaaattcc tgcacgacat gctggcgga gaactccgct attccgtgat ccgagaagtg      420
ctgccaacgc ggcggggcacg aacgttcggc ctcgaggctg aggaactgca caccctcgtc     480
gccgaagggg tccttgtaca caac                                           504
```

<210> 4

<211> 168

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
TOPO modified intein sequence from SEQ ID NO: 2

<400> 4

```
Ala Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr
1              5              10              15
```

```
His Arg Ile Glu Asp Val Val Gly Gly Arg Lys Pro Ile His Val Val
20              25              30
```

```
Ala Ala Ala Lys Asp Gly Thr Leu His Ala Arg Pro Val Val Ser Trp
35              40              45
```

```
Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly
50              55              60
```

```
Ala Ile Leu Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65              70              75              80
```

```
Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gln Pro
```

Arg Arg Phe Asp Gly Phe Gly Asp Ser Ala Pro Ile Pro Ala Arg Val  
 100 105 110

Gln Ala Leu Ala Asp Ala Leu Asp Asp Lys Phe Leu His Asp Met Leu  
 115 120 125

Ala Glu Glu Leu Arg Tyr Ser Val Ile Arg Glu Val Leu Pro Thr Arg  
 130 135 140

Arg Ala Arg Thr Phe Gly Leu Glu Val Glu Glu Leu His Thr Leu Val  
 145 150 155 160

Ala Glu Gly Val Leu Val His Asn  
 165

<210> 5

<211> 531

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

GATEWAY modified intein sequence from SEQ ID NO: 1

<400> 5

```

gccctcgcag agggcactcg gatcttcgat ccggtcaccg gtacaacgca tcgcatcgag      60
gatgttgctg gtgggcgcaa gcctattcat gtcgtggctg ctgccaagga cggaacgctg      120
catgcgcggc ccgtggtgtc ctggttcgac cagggaacgc gggatgtgat cggggttcgg      180
atcgccggtg ggcctatcct gtgggcgaca cccgatcaca aggtgctgac agagtacggc      240
tggcgtgccg ccgggggaact ccgcaaggga gacaggggtg cgcaaccgag acgcttcgat      300
ggattcggtg acagtgcgcc gattccgaca agtttgtaca aaaaagcagg cagcgcgcgc      360
gtgcaggcgc tcgcggatgc cctggatgac aaattcctgc acgacatgct ggcggaagaa      420
ctccgctatt ccgtgatccg agaagtgtg ccaacgcggc gggcacgaac gttcggcctc      480
gaggtcgcagg aactgcacac cctcgtcgcc gaaggggttg ttgtacacaa c              531

```

<210> 6

<211> 177

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

GATEWAY modified intein sequence from SEQ ID NO: 2

<400> 6

Ala Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr  
1 5 10 15

His Arg Ile Glu Asp Val Val Gly Gly Arg Lys Pro Ile His Val Val  
20 25 30

Ala Ala Ala Lys Asp Gly Thr Leu His Ala Arg Pro Val Val Ser Trp  
35 40 45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly  
50 55 60

Ala Ile Leu Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly  
65 70 75 80

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gln Pro  
85 90 95

Arg Arg Phe Asp Gly Phe Gly Asp Ser Ala Pro Ile Pro Thr Ser Leu  
100 105 110

Tyr Lys Lys Ala Gly Ser Ala Arg Val Gln Ala Leu Ala Asp Ala Leu  
115 120 125

Asp Asp Lys Phe Leu His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser  
130 135 140

Val Ile Arg Glu Val Leu Pro Thr Arg Arg Ala Arg Thr Phe Gly Leu  
145 150 155 160

Glu Val Glu Glu Leu His Thr Leu Val Ala Glu Gly Val Val Val His  
165 170 175

Asn

<210> 7

<211> 525

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
TOPO and GATEWAY modified intein sequence from  
SEQ ID NO: 1

<400> 7

```
gcctctgcag agggcactcg gatcttcgat ccggtcaccg gtacaacgca tcgcatcgag      60
gatgttgctg gtggggcgcaa gcctattcat gtcgtggctg ctgccaaagga cggaacgctg    120
catgcgcggc ccgtggtgtc ctggttcgac cagggaacgc gggatgtgat cggggtgcgg      180
atcgccggtg gcgccatcct gtggggcgaca cccgatcaca aggtgctgac agagtacggc      240
tggcgtgccg ccgggggaact ccgcaaggga gacagggtgg cgcaaccgag acgcttcgat      300
ggattcgggtg acagtgcgcc gattccgaca agtttgtaga aaaaagcagg cagcgcgcgc      360
gtgcaggcgc tcgcggatgc cctggatgac aaattcctgc acgacatgct ggcggaagaa      420
ctccgctatt ccgtgatccg agaagtgtg ccaacgcggc gggcacgaac gttcggcctc      480
gaggaactgc acaccctcgt cgccgaaggg gtccttgtag acaac                        525
```

<210> 8

<211> 177

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
TOPO and GATEWAY modified intein sequence from  
SEQ ID NO: 2

<400> 8

```
Ala Leu Ala Glu Gly Thr Arg Ile Phe Asp Pro Val Thr Gly Thr Thr
1              5              10             15

His Arg Ile Glu Asp Val Val Gly Gly Arg Lys Pro Ile His Val Val
20             25             30

Ala Ala Ala Lys Asp Gly Thr Leu His Ala Arg Pro Val Val Ser Trp
35             40             45

Phe Asp Gln Gly Thr Arg Asp Val Ile Gly Leu Arg Ile Ala Gly Gly
50             55             60

Ala Ile Leu Trp Ala Thr Pro Asp His Lys Val Leu Thr Glu Tyr Gly
65             70             75             80
```

Trp Arg Ala Ala Gly Glu Leu Arg Lys Gly Asp Arg Val Ala Gln Pro  
85 90 95

Arg Arg Phe Asp Gly Phe Gly Asp Ser Ala Pro Ile Pro Thr Ser Leu  
100 105 110

Tyr Lys Lys Ala Gly Ser Ala Arg Val Gln Ala Leu Ala Asp Ala Leu  
115 120 125

Asp Asp Lys Phe Leu His Asp Met Leu Ala Glu Glu Leu Arg Tyr Ser  
130 135 140

Val Ile Arg Glu Val Leu Pro Thr Arg Arg Ala Arg Thr Phe Gly Leu  
145 150 155 160

Glu Val Glu Glu Leu His Thr Leu Val Ala Glu Gly Val Leu Val His  
165 170 175

Asn

<210> 9

<211> 6262

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
pET-GWMIT sequence

<220>

<221> modified\_base

<222> (817)..(820)

<223> a, c, g, t, unknown or other

<220>

<221> modified\_base

<222> (822)..(822)

<223> Inosine

<220>

<221> modified\_base

<222> (823)..(826)

<223> a, c, g, t, unknown or other

<220>

<221> modified\_base

<222> (827)..(828)

<223> Inosine

<220>

<221> modified\_base

<222> (829)..(841)

<223> a, c, g, t, unknown or other

<400> 9

caaggagatg ggcgccaaca gtcccccggc cacgggggect gccaccatac ccacgccgaa	60
acaagcgctc atgagcccga agtggcgagc ccgatcttcc ccatcgggtga tgtcggcgat	120
ataggcgcca gcaaccgcac ctgtggcgcc ggtgatgccg gccacgatgc gtccggcgta	180
gaggatcgag atctcgatcc cgcgaaatta atacgactca ctatagggga attgtgagcg	240
gataacaatt cccctctaga aataattttg tttaacttta agaaggaatt cgccctcgca	300
gagggcactc ggatcttcga tccggtcacc ggtacaacgc atcgcatcga ggatgttgtc	360
ggtgggcgca agcctattca tgtcgtggct gctgccaaag acggaacgct gcatgcgcgg	420
cccggtgtgt cctgggtcga ccagggaacg cgggatgtga tcgggttgcg gatcgccggt	480
ggcgccatcc tgtgggcgac acccgatcac aaggtgctga cagagtacgg ctggcgtgcc	540
gccggggaac tccgcaagg agacaggggtg gcgcaaccgc gacgcttcga tggattcgggt	600
gacagtgcgc cgattccgac aagtttgtac aaaaaagcag gcagcgcgcg cgtgcaggcg	660
ctcgcggatg ccttgatga caaatctctg cacgacatgc tggcggaaga actccgctat	720
tccgtgatcc gagaagtgtc gccaacgcgg cgggcacgaa cgttcggcct cgaggaactg	780
cacaccctcg tcgcgaagg ggttcttgta cacaacnnnn mnnnnnnnnn nnnnnnnnnn	840
nggtaagcct atccctaacc ctctctcgg tctcgattct acgcgtaccg gtcacatca	900
ccatcaccat tgagtttgat ccggtgcta acaaagccc aaaggaagct gagttggctg	960
ctgccaccgc tgagcaataa ctagcataac cccttggggc ctctaaacgg gtcttgaggg	1020
gttttttgtc gaaaggagga actatatccg gatatcccgc aagaggcccg gcagtaccgg	1080
cataaccaag cctatgccta cagcatccag ggtgacggtg ccgaggatga cgatgagcgc	1140
attgttagat ttcatacacg gtgcctgact gcgttagcaa tttaactgtg ataaactacc	1200
gcattaaagc ttatcgatga taagctgtca aacatgagaa ttaattcttg aagacgaaag	1260
ggcctcgtga tacgcctatt tttatagggt aatgtcatga taataatgggt ttcttagacg	1320
tcagggtggca cttttcgggg aaatgtgcgc ggaacccta tttgtttatt tttctaaata	1380
cattcaaata tgtatccgct catgagacaa taaccctgat aaatgcttca ataatttga	1440
aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc ttattccctt ttttgcggca	1500

ttttgccttc ctgtttttgc tcaccagaa acgctggtga aagtaaaaga tgctgaagat	1560
cagttgggtg cagcagtggtg ttacatcgaa ctggatctca acagcggtaa gatccttgag	1620
agttttcgcc ccgaagaacg ttttccaatg atgagcactt ttaaagttct gctatgtggc	1680
gcggtattat cccgtgttga cgccgggcaa gagcaactcg gtcgccgat acactattct	1740
cagaatgact tggttgagta ctcaccagtc acagaaaagc atcttacgga tggcatgaca	1800
gtaagagaat tatgcagtgc tgccataacc atgagtgata aactgcggc caacttactt	1860
ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaacat gggggatcat	1920
gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt	1980
gacaccacga tgctgcagc aatggcaaca acgttgcgca aactattaac tggcgaacta	2040
cttactctag cttcccggca acaattaata gactggatgg aggcggataa agttgcagga	2100
ccacttctgc gctcgccct tccggctggc tggtttattg ctgataaatc tggagccggt	2160
gagcgtgggt ctgcgggtat cattgcagca ctggggccag atggtaaagc ctcccgtatc	2220
gtagttatct acacgacggg gagtcaggca actatggatg aacgaaatag acagatcgct	2280
gagatagggtg cctcactgat taagcattgg taactgtcag accaagttta ctcatatata	2340
ctttagattg atttaaaact tcatttttaa tttaaaagga tctaggtgaa gatccttttt	2400
gataatctca tgaccaaaat cccttaacgt gagttttcgt tccactgagc gtcagacccc	2460
gtagaaaaga tcaaaggatc ttcttgagat cctttttttc tgccgtaat ctgctgcttg	2520
caaacaaaaa aaccacgct accagcgggtg gtttgtttgc cggatcaaga gctaccaact	2580
ctttttccga aggtaactgg cttcagcaga gcgcagatac caaatactgt ccttctagt	2640
tagccgtagt taggccacca cttcaagaac tctgtagcac cgctacata cctcgctctg	2700
ctaatectgt taccagtggc tgetgccagt ggcgataagt cgtgtcttac cgggttggac	2760
tcaagacgat agttaccgga taaggcgag cggtcgggct gaacgggggg ttcgtgcaca	2820
cagcccagct tggagcgaac gacctacacc gaactgagat acctacagcg tgagctatga	2880
gaaagcgcca cgcttcccga agggagaaaag gcggacaggt atccgtaag cggcagggtc	2940
ggaacaggag agcgcacgag ggagcttcca gggggaaacg cctggtatct ttatagtcct	3000
gtcgggtttc gccacctctg acttgagcgt cgatttttgt gatgctcgtc aggggggcgg	3060
agcctatgga aaaacgccag caacgcggcc tttttacggg tccctggcctt ttgctggcct	3120
tttgctcaca tgttctttcc tgcgttatcc cctgattctg tggataaccg tattaccgcc	3180
tttgagtgag ctgataccgc tcgccgcagc cgaacgaccg agcgcagcga gtcagtgagc	3240

gaggaagcgg aagagcgcct gatgcggtat tttctcctta cgcattctgtg cggatatttca	3300
caccgcaatg gtgcactctc agtacaatct gctctgatgc cgcatagtta agccagtata	3360
cactccgcta tcgctacgtg actgggtcat ggctgcgccc cgacacccgc caacacccgc	3420
tgacgcgccc tgacgggctt gtctgctccc ggcatccgct tacagacaag ctgtgaccgt	3480
ctccgggagc tgcattgtgc agaggttttc accgtcatca ccgaaacgcg cgaggcagct	3540
gcggtaaagc tcatcagcgt ggctcgtgaag cgattcacag atgtctgcct gttcatccgc	3600
gtccagctcg ttgagtttct ccagaagcgt taatgtctgg cttctgataa agcgggccat	3660
gttaagggcg gttttttcct gtttgggtcac tgatgcctcc gtgtaagggg gatttctgtt	3720
catgggggta atgataccga tgaaacgaga gaggatgctc acgatacggg ttactgatga	3780
tgaacatgcc cggttactgg aacgttgtga gggtaaacaa ctggcggtat ggatgcggcg	3840
ggaccagaga aaaatcactc agggccaatg ccagcgcttc gttaatacag atgtaggtgt	3900
tccacagggc agccagcagc atcctgcgat gcagatccgg aacataatgg tgcagggcgc	3960
tgacttccgc gtttccagac tttacgaaac acggaaaccg aagaccattc atgttgttgc	4020
tcaggctcga gacgttttgc agcagcagtc gcttcacgtt cgctcgcgta tcggtgattc	4080
attctgctaa ccagtaaggc aaccccgcca gcctagccgg gtctcaacg acaggagcac	4140
gatcatgcgc acccgtggcc aggacccaac gctgcccag atgcgccgcg tgcggctgct	4200
ggagatggcg gacgcgatgg atatgttctg ccaagggttg gtttgcgcat tcacagttct	4260
ccgcaagaat tgattggctc caattcttgg agtgggtgaat ccgttagcga ggtgccgccg	4320
gcttccattc aggtcgaggt ggcccggctc catgcaccgc gacgcaacgc ggggaggcag	4380
acaaggtata gggcggcgc tacaatccat gccaaaccgt tccatgtget cgccgaggcg	4440
gcataaatcg ccgtgacgat cagcgggtcca atgatcgaag ttaggtggt aagagccgcg	4500
agcgatcctt gaagctgtcc ctgatggctg tcatctacct gcctggacag catggcctgc	4560
aacgcgggca tcccgatgcc gccggaagcg agaagaatca taatggggaa ggccatccag	4620
cctcgcgtcg cgaacgccag caagacgtag ccagcgcgt cggccgccat gccggcgata	4680
atggcctgct tctcgccgaa acgtttggtg gcgggaccag tgacgaaggc ttgagcgagg	4740
gcgtgcaaga ttccgaatac cgcaagcgac aggccgatca tcgtcgcgct ccagcgaaag	4800
cggctctcgc cgaaaatgac ccagagcgct gccggcacct gtctacgag ttgcatgata	4860
aagaagacag tcataagtgc ggcgacgata gtcattgccc gcgccaccg gaaggagctg	4920

actggggtga aggtctctcaa gggcatcggt cgagatcccg gtgcctaatag agtgagctaa	4980
cttacattaa ttgcgttgcg ctcactgccc gctttccagt cgggaaacct gtcgtgccag	5040
ctgcattaat gaatcggcca acgcgcgggg agaggcggtt tgcgtattgg gcgccagggt	5100
ggtttttctt ttcaccagtg agacgggcaa cagctgattg cccttcaccg cctggccctg	5160
agagagttgc agcaagecgt ccacgctggt ttgccccagc ag	